REMARKS

This Amendment is submitted in response to the Office Action dated December 29, 2005, having a shortened statutory period set to expire March 29, 2006. Proposed amendments are submitted for Claims 8, 10 and 13, Claims 15-20 are added, and Claims 1-7, 9 and 14 are cancelled. Upon entry of the proposed amendments, Claims 8, 10-13 and 15-20 will be pending.

Applicant's undersigned representative appreciates the time and courtesy extended by the Examiner during a teleconference held on March 27, 2006. No formal agreement was reached during this teleconference.

Rejections under 35 U.S.C. §102

In paragraph 5 of the present Office Action, the Examiner has rejected Claims 1-14 under 35 U.S.C. § 102(b) as being anticipated by *Ishikawa et al.* (U.S. Patent No. 5,848,407 – "*Ishikawa*"). In light of the pending amendments, Applicants respectfully traverse these rejections.

With regards to amended Claim 8, the cited art does not teach or suggest "conducting a structure analysis of said web page, wherein the structure analysis includes the steps of: reading an HTML document of a web page as an analyzing object, conducting a temporary block analysis based on a description of HTML tags of the HTML document (as supported in the present specification at paragraph [0075]), using the HTML tags to temporarily divide the HTML document into blocks, and identifying unnecessary information elements in the HTML document, wherein the unnecessary information elements are plural information elements that include an OBJECT_IMAGE having a same Uniform Resource Locator (URL), (as supported at paragraph [0082]) wherein the OBJECT_IMAGE describes a type of media used to display the HTML document (as supported at paragraph [0059])."

With reference to Claim 10, the cited art does not teach or suggest "a process of calculating a degree of significance of a web site linking from said web page, based on the result

of said structure analysis stored in said storage device, wherein scores used to calculate a degree of significance are calculated based on information elements added to anchors through an analysis of the web page (see paragraph [0115] of the present specification for support), wherein the analysis of the web page includes the steps of:

reading an HTML document of a web page as an analyzing object (supported at paragraph [0075]),

conducting a temporary block analysis based on a description of HTML tags of the HTML document (supported at paragraph [0075]),

using the HTML tags to temporarily divide the HTML document into blocks (supported at paragraph [0075]), and

identifying unnecessary information elements in the HTML document, wherein the unnecessary information elements are plural information elements that include an OBJECT_IMAGE having a same Uniform Resource Locator (URL),(supported at paragraph [0082]) wherein the OBJECT_IMAGE describes a type of media used to display the HTML document (supported at paragraph [0059])."

Claim 13 has been amended to include the features previously claimed in dependent Claim 14, plus additional features. As amended, Claim 13 is patentable since the cited art does not teach or suggest "a second process of reading the blocked structural data of said HTML document from said memory, updating block structures of said HTML document by associating the information elements that are mutually relevant in terms of a meaning, and storing the updated structural data into the memory, wherein said second process includes the steps of:

identifying an unnecessary information element in terms [[of a purpose]] of a document structure analysis, wherein an information element is deemed to be unnecessary if the information element includes an OBJECT IMAGE that includes a Uniform Resource Locator (URL) that has been used by another information element in the HTML document (as supported by paragraph [0082] of the present specification), wherein the OBJECT IMAGE describes a type of media used to display the HTML document (see paragraph [0059] for support);

merging said information elements or dividing a block based on contents of said information elements; and

merging the block structures based on information contained in each block.

With regards to newly added Claim 15, the cited art does not teach or suggest a method that comprises "reading an HTML document of a web page as an analyzing object (as supported by paragraph [0075] of the present specification);

conducting a temporary block analysis based on a description of HTML tags of the HTML document (supported by paragraph [0075]);

using the HTML tags to temporarily divide the HTML document into blocks (supported by paragraph [0075]; and

identifying unnecessary information elements in the HTML document, wherein the unnecessary information elements include plural information elements that include an OBJECT_IMAGE having a same Uniform Resource Locator (URL) (supported by paragraph [0082]), wherein the OBJECT_IMAGE describes a type of media used to display the HTML document (supported by paragraph [0059]);

deleting any block in the HTML document that is deemed to be structurally meaningless, wherein a block is deemed to be structurally meaningless if that block has only unnecessary information elements (supported by paragraph [0087]); and ·

merging relevant information elements in a same block into one composite element (supported by paragraph [0088]).

With regards to newly added dependent Claim 16, the cited art does not teach or suggest "wherein the unnecessary information elements include OBJECT_ANCHORS that have a same title (as supported by paragraph [0082] of the present specification), wherein an OBJECT_ANCHOR describes a correlation between the HTML document and elements in another web page (as supported by paragraph [0105]).

With regards to newly added dependent Claim 17, the cited art does not teach or suggest "wherein the unnecessary information elements include OBJECT_TEXT_BLOCKS that have a same description of text in a block," as supported in the present specification at paragraph [0082].

With regards to newly added dependent Claim 18, the cited art does not teach or suggest "wherein the relevant information elements that are merged are from a group that includes the OBJECT_IMAGE, OBJECT_ANCHOR and OBJECT_TEXT_BLOCKS," as supported in the present specification at paragraph [0089]).

With regards to newly added independent Claim 19, the cited art does not teach or suggest a method for eliminating ambiguity of a specified topic being searched during a web crawling, which is supported in the present specification at paragraph [0189], and includes the steps of:

"presenting relevant keywords to a user during web crawling, wherein the relevant keywords describe multiple attributes of a term that has an ambiguous meaning, and wherein the user is afforded an ability to specify keywords that have a minus degree of significance to a meaning intended by the user for web crawling; and

narrowing down crawling objects by eliminating user-specified keywords that have a minus degree of significance, thereby eliminating ambiguity of a term being searched."

With regards to newly added independent Claim 20, the cited art does not teach or suggest a web crawler as illustrated in Figure 2 of the present disclosure, and includes the elements:

"an initial site acquiring section, wherein the initial site acquiring section specifies a Uniform Resource Locator (URL) of a home page of a specific web site from which information is to be collected, and wherein initial web sites to be searched are obtained through the use of keywords in a search engine (supported by paragraph [0050] of the present specification), and wherein the initial web sites represent a set of web sites that are initially set for web crawling (supported by paragraph [0047]);

a document structure analysis section for performing document structure analysis for a web page of the initial sites (supported by paragraph [0049]), wherein the document structure analysis includes the steps of:

reading an HTML document of a web page as an analyzing object,

conducting a temporary block analysis based on a description of HTML tags of the HTML document (supported by paragraph [0075]),

using the HTML tags to temporarily divide the HTML document into blocks, and

identifying unnecessary information elements in the HTML document, wherein

the unnecessary information elements are plural information elements that include an

OBJECT_IMAGE having a same Uniform Resource Locator (URL) (supported by

paragraph [0082]), wherein the OBJECT_IMAGE describes a type of media used to display the

HTML document (see paragraph [0059]);

a significance calculating section for calculating degrees of significance of web sites that

are acquired by web crawling, wherein the degrees of significance are based on a result of the

document structure analysis performed by the document structure analysis section (see paragraph

[0049]), and wherein calculating degrees of significance extends a Fish-Search crawling

technique by basing the calculating on strategies specified by a user and information elements

added to anchors through the document structure analysis performed by the document structure

analysis section, and wherein objects of crawling are dynamically determined depending on the

degrees of significance (see paragraph [0115]); and

a crawling executing section for executing a process of acquiring web sites by crawling

based on the results of the degrees of significance calculated by the significance calculating

section (see paragraph [0049])."

CONCLUSION

As the cited art does not teach or suggest all of the presently claimed features, Applicants now respectfully request a Notice of Allowance for all pending claims.

No extension of time for this response is believed to be necessary. However, in the event an extension of time is required, that extension of time is hereby requested. Please charge any fee associated with an extension of time as well as any other fee necessary to further the prosecution of this application to **IBM CORPORATION DEPOSIT ACCOUNT No. 09-0461.**

Respectfully submitted,

James E. Boice
Registration No. 44,545
DILLON & YUDELL LLP
8911 North Capital of Texas Highway
Suite 2110
Austin, Texas 78759
512.343.6116

ATTORNEY FOR APPLICANT(S)

JP920020109US1 – Amendment A   -12-   10/621,474